



OPEN

# Genetic and environmental links between traits of cocoa beans and pods clarify the phenotyping processes to be implemented

Fabien Doaré<sup>1</sup>, Fabienne Ribeyre<sup>2</sup> & Christian Cilas<sup>2</sup>✉

The average weight of cocoa beans is not generally taken into account during breeding processes, although it is a trait of interest. Several studies indicate that the weight of the beans has a high heritability in *Theobroma cacao*. However, the values obtained from different countries for the same clone often vary. In this study, we analyzed the effect of different factors on the weight of the beans. Apart from the clone effect, three main factors had an impact: i) the number of beans per pod: a good filling of the pod with beans tended to limit the weight of the beans, ii) the position of the beans in the pod: beans in the apical part of the pod were significantly lighter than the others and iii) the longer the duration of the fructification cycle the heavier the beans were (positive genetics correlation). These results lead us to propose protocols aimed at normalizing the phenotypic values of the genetic material. To obtain a reliable estimate of the bean weight, the following is proposed: either to use beans obtained from manual pollination to saturate the pods with beans, or to systematically use the number beans in the pods as a covariable.

The cacao tree (*T. cacao*) is a small, evergreen tree native to the Amazonian forest<sup>1</sup>. This tree of the Malvaceae family (formerly Sterculiaceae) is grown for its beans contained in the fruits. The cacao tree's growth is indeterminate and its fruit production can vary from 20 to over 100 years, with fruit growing on both trunk and branches. The fruits, known as cacao pods, arise from the pollination of flowers grouped in flower cushions<sup>2</sup>. The pods contain beans, which are extracted from the pod and fermented with surrounding mucilage, then dried to produce fermented dried cocoa, the raw material used to make chocolate. The production of commercial cocoa beans from one tree over a given period depends on the number of pods produced, the number of seeds per pod and the mean seed weight<sup>3</sup>.

The main goals of cocoa breeding are to increase the cocoa production per tree and to improve the cocoa quality. To increase the production per tree implies: i) improved resistance to the pest and diseases of cocoa, ii) increased pods produced per tree, iii) increased number of beans per pod and iv) increased average bean weight.

In fact, the production  $Y_i$  of a tree  $i$  between date  $t$  and date  $t + s$ , ( $Y_i[t, t + s]$ ) can be written as:

$$Y_i[t, t + s] = \text{NoPods}_i[t, t + s] \times \text{NoSPi}[t, t + s] \times \text{SWi}[t, t + s]$$

where:

$\text{NoPods}_i[t, t + s]$ : number of healthy pods produced by tree  $i$  between date  $t$  and date  $t + s$ .

$\text{NoSPi}[t, t + s]$ : mean number of seeds per pod for tree  $i$  between date  $t$  and date  $t + s$ .

$\text{SWi}[t, t + s]$ : mean seed weight for tree  $i$  between date  $t$  and date  $t + s$ .

The heritability of the number of healthy pods produced per tree is relatively weak ( $<0.4$ ), which indicates that this trait is mostly dependent on environmental factors<sup>4–5</sup>. Actually, the trait is the result of several elementary traits or processes: flowering, pollination, fruit-setting, cherelle wilt, diseases and insect attacks<sup>6–8</sup>. In addition, these elementary traits, with generally low heritabilities, are instead dependent on environmental conditions, like soil and climate.

<sup>1</sup>CIRAD, UPR Bioagresseurs, F- 97387 Kourou, Guyane, France. Bioagresseurs, Univ Montpellier, CIRAD, Montpellier, Kourou, France. <sup>2</sup>CIRAD, UPR Bioagresseurs, F-34398 Montpellier, France. Bioagresseurs, Univ Montpellier, CIRAD, Montpellier, France. ✉e-mail: [christian.cilas@cirad.fr](mailto:christian.cilas@cirad.fr)

The number of seeds per pod and the mean seed weight are traits that should be also improved to increase production per tree. The number of seeds per pod depends on several factors, including the number of ovules per ovary, the fertility of the ovules, which varies according to the self-compatibility or self-incompatibility of the cacao genotype, and natural pollination conditions<sup>9</sup>. Mean seed weight is an important trait with high heritability<sup>10</sup>, a large bean size as well as uniformity of the bean size being the characteristics of interest for the chocolate industry. The average bean weight is therefore an essential trait to take into account in the genetic improvement of cocoa.

Several studies have revealed that the mean seed weight showed a normal distribution per clone and a high heritability ( $\geq 0.5$ )<sup>10–12</sup>, but the values obtained in different studies remain quite variable for the same clones. We wanted to understand how this trait varied according to several factors, including the filling of pods with beans, and the position of the beans in the pods. Observations on several clones in the CIRAD collection in French Guiana were conducted to answer these questions: i) what is the impact of filling the pods with beans on the weight of the beans? ii) is there a heterogeneity in the size of the beans in the pods (between the peduncular, medial and apical pod sectors)? what is the link with other factors like the duration of the fruiting cycle (from pollination to maturation)? This is the first time (as far as we are aware) that these criteria (number of beans per pod, position of the bean in the pod, duration of the fruiting cycle) have been taken into account to explain the bean weight, in addition to the genetic effect with the study of several clones.

## Material and methods

**Plant material.** Seven clones situated in the Pointe-Combi experimental station of CIRAD in French Guiana were used for this study (ELP 35 A, GF 23, CCN 51, ICS 60, IMC 97, NA 79, PA 121).

ELP 35 A: clone collected in the area of Crique Euleupoussing in the French Guiana forest (Guiana group)

GF 23: clone collected in French Guiana coast (Amelonado group),

CCN 51: clone coming from a cross of the hybrid (IMC 67 x ICS 95) with an Ecuadorian cultivar locally known as “Canellos” (Hybrid),

ICS 60: clone coming from selections in Trinidadian farms, (Criollo group),

IMC 97: clone collected by Pound in Iquitos area (Peru), (Iquitos group),

NA 79: clone collected by Pound in Nanay area (Peru), (Nanay group),

PA 121: clone collected by Pound in Parinari area (Peru), (Marañon group).

These clones were chosen because of i) their good representativity of the genetic diversity of *T. cacao*<sup>13</sup>, ii) their availability of flowers and pods, and iii) their good physiological condition in the Pointe-Combi experimental station. Several trees per clone were available for the study, between 2 and 16 trees per clone, depending on the clone.

**Observations.** The mature pods were harvested and taken to the CIRAD laboratory in Kourou (French Guiana) between February 2016 and March 2019. On the same day of harvesting, the following were measured: weight of the whole pod, weight of the wet beans and weight of the cortex. The beans of each pod were counted and separated into 3 sectors of equivalent size (Fig. 1): peduncular sector, median sector and apical sector. The beans from each sector were weighed and the average weight of a bean determined for each sector. In addition, fruiting cycle durations (FCD) – from pollination to pod maturation – were measured on a sample of manual pollination pods for each of the seven clones.

**Data analysis.** For the number of beans per pod, the pod weights and the total weight of the beans, one way analyses of variance were performed with the factor “clone”. For pod weights and overall bean weight covariance analyses were also performed using the number of beans per pod as a covariate.

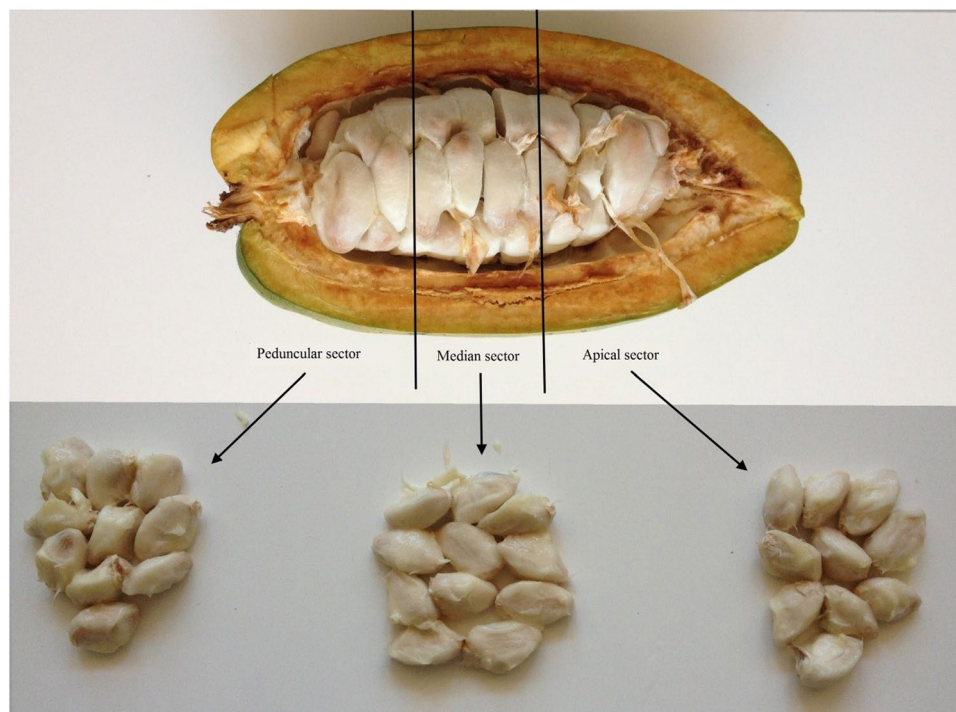
For the bean weight per sector, two way analyses of variance were performed (clone and sector) as well as covariance analyses using the number of beans per pod as covariate.

The broad sense heritability for these different traits was estimated, along with the associated confidence intervals. The confidence intervals were estimated by the Wald method<sup>14</sup>. Genetic and environmental correlations between the several pod and bean traits on one hand, and the duration of the fruiting cycle on the other hand were estimated using a multivariate general linear model (for the pods with the duration of the fruiting cycle data). Data analyses were performed with Proc GLM and Proc Mixed (SAS system)<sup>15</sup>.

## Results

**Comparison of clones and heritability values.** The number of beans per pod had a moderate heritability (0.27). The other pods and beans traits were very different depending on the clones, with very strong heritabilities (Table 1), especially for the average weight of a bean (0.70).

The number of beans per pod was significantly different among clones, from 24.7 beans per pod for PA 121 to 44.9 beans per pod for IMC 97 (Table 1). The clone IMC 97 had a significantly higher number of beans per pod than all the other clones tested. Three other homogeneous groups of clones were determined. ICS 60 and GF 23 had a higher number of beans than NA 79. This one had a higher number of beans than ELP 34 A and PA 121, with CCN 51 being between the two last groups. The differences among clones were significant for all the traits studied. IMC 97 also had higher values for total pod weight, cortex weight and total bean weight per pod (Table 1). On the other hand, CCN 51 clone had the highest average bean weight (Table 1). Indeed, for the average bean weight, mean values per clone ranged from 2.24 for GF 23 to 5.64 for CCN 51. Five groups were identified: CCN51 had the highest mean bean weight followed by ICS 60 that had a higher weight of a bean than IMC 97. GF23 had a significantly lower mean weight bean than all other clones.



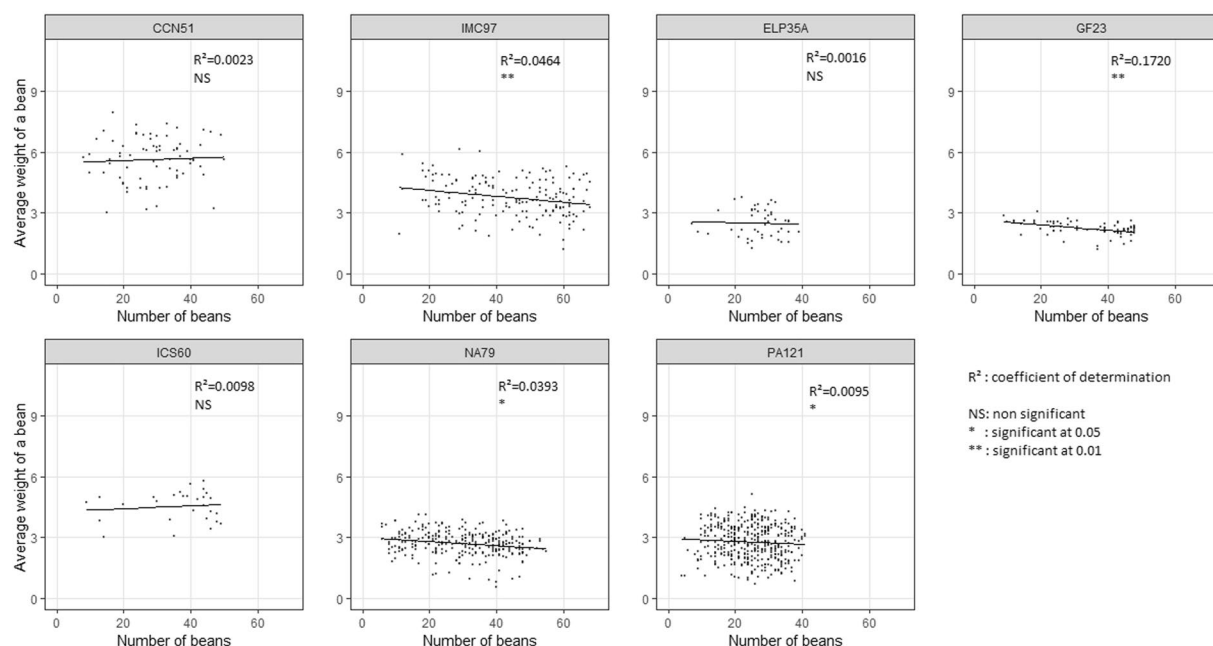
**Figure 1.** Study of the seeds according to the three sectors of the pod: Peduncular sector, Median sector, Apical sector.

Clone	NBeans	PodW	CortexW	BeansW	BeanWm	BeanWm Cov: NBeans
IMC 97	44.9 a*	714.8 a	548.8 a	166.1 a	3.80 c	3.95 c
ICS 60	36.4 b	592.5 c	426.2 c	166.3 a	4.54 b	4.61 b
GF 23	34.0 b	312.2 d	238.0 d	74.2 b	2.24 e	2.27 e
NA 79	30.3 c	339.4 d	259.1 d	80.3 b	2.70 d	2.70 d
CCN 51	29.0 cd	659.9 b	495.7 b	164.2 a	5.64 a	5.63 a
ELP 35 A	27.2 d	351.1 d	283.1 d	68.0 b	2.51 d	2.47 d
PA 121	24.7 d	310.2 d	241.6 d	68.5 b	2.79 d	2.72 d
**h <sup>2</sup>	0.27	0.54	0.49	0.59	0.70	0.72
***CI	[0.04; 0.50]	[0.25; 0.82]	[0.21; 0.78]	[0.31; 0.86]	[0.47; 0.94]	[0.48; 0.95]

**Table 1.** Comparison of clones for the number of beans per pod (*NBeans*) for the weight of a pod (*PodW*) in g, for the weight of the cortex (*CortexW*) in g, for the weight of beans per pod (*BeansW*) in g, for the average weight of a bean (*BeanWm*) in g, and multiple mean comparison test (Newman-Keuls 5%)\*. For the last column, average weight of a bean adjusted to the covariate: number of beans per pod. Broad sense heritabilities ( $h^2$ ) of the traits with confidence interval at 5% (CI). \* means followed by the same letter are not significantly different at the 5% threshold by the Newman & Keuls test \*\*h<sup>2</sup>: broad sense heritability \*\*\* CI: confidence intervals estimated by the Wald method.

Source	*DF	**Type III SS	Root mean square	*** F value	****Pr> F
Clone	6	2279.5	379.92	610.35	<0.0001
Sector	2	46.1	23.06	37.05	<0.0001
Clone*Sector	12	18.8	1.56	2.51	0.003
NBeans	1	31.9	31.91	51.27	<0.0001

**Table 2.** General linear model for the mean weight of a bean according to clone and sector factors, and the covariate “number of beans per pod”. \* degrees of freedom. \*\* type III sums of squares (all effects are adjusted to all others). \*\*\* value of the Fisher test \*\*\*\* the significance value for the test.



**Figure 2.** Regression between the average weight of a bean and the filling of the pods with beans for each of the 7 clones.

Clone	BeanWm	Sector	BeanWm
CCN 51	5.62 a*	Pedunc.	3.62 a
ICS 60	4.60 b	Median	3.56 a
IMC 97	3.88 c	Apic.	3.22 b
Pa 121	2.79 d		
Na 79	2.70 d		
ELP 35 A	2.48 e		
GF 23	2.25 f		

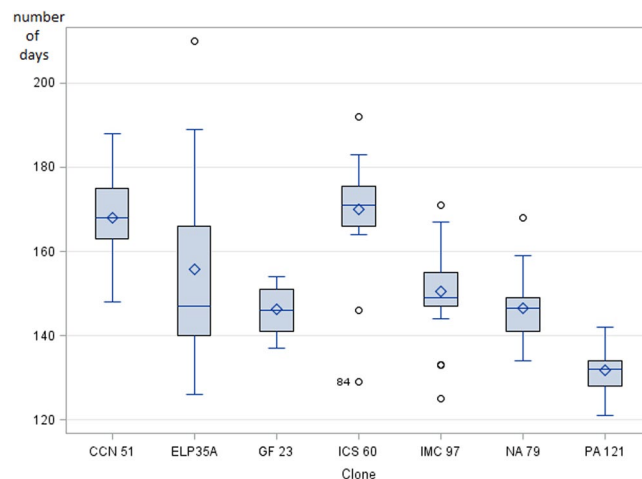
**Table 3.** Comparison of clones means and sectors means, with cov: *NBeans*. \* means followed by the same letter are not significantly different at the 5% threshold by the Newman & Keuls test.

For the average bean weight, the covariance analysis indicated that there was an effect of the covariate: “number of beans per pod” (Table 2). Pods better filled with beans have lighter beans on average for most clones (Fig. 2). However, the means adjusted to this covariate were in the same order for the 7 clones considered (Table 1). To verify if the effect of the covariate was the same for each of the 7 clones, a regression of the average bean weight on the number of beans per pod was performed for each of the 7 clones (Fig. 2), and the regressions were significant for only 4 clones (IMC 97, GF 23, NA 79 and PA 121).

**Sources of bean weight variability.** The effect of the sectors (peduncle, median, apical) on the average bean weight was then taken into account. The clone and sector effects were very significant and the interaction between these two factors was also significant, but with a lesser effect (Table 2). The effect of the sector therefore existed whatever the clone.

Comparisons of clone and sector means were then made using this model (Table 3). The clone means were similar to those previously estimated (for instance, for CCN 51, the estimated bean weight was 5.64 g, 5.63 g taking into account the pod filling and 5.62 g taking into account also the pod sector), and therefore the clone ranking was not changed. For other clones, the differences between the several assessments were more important. There was a strong difference between the 3 sectors considered with a significantly lower bean weight mean in the apical pod region (reduction of 15% on average in comparison to the two other sectors), (Table 3).

Fruit cycle durations (FCD) were also observed for each of the 7 clones studied and significant differences were detected, from 132 days for PA 121 to 170 days for ICS 60 (Fig. 3). An important variability was observed for the ELP 35 A clone; indeed, this clone did not show very clear signs of maturation with pods that remained partially green. FCD is genetically correlated to the pods and beans traits, particularly the beans weight (Table 4), with genetic correlations of 0.81 and 0.80 (with *BeansW* and *BeansWm* respectively). The longer the fruiting cycles of the clones, the heavier were the beans.



**Figure 3.** Box plot of fruiting cycle durations (FCD) of the seven clones (unit: number of days).

FCD	NBeans	PodW	CortexW	BeansW	BeanWm
G correlation	0.46	0.69	0.63	0.81	0.80
R correlation	−0.05	0.01	0.03	−0.05	−0.04

**Table 4.** Genetic correlations and residual correlations between the studied traits and fruit cycle duration (FCD).

## Discussion

There is a strong genetic effect on the weight of pods and beans, which confirms previous studies<sup>3,10,16</sup>. The effect of pod filling with beans on the average weight of a bean is highlighted; on average, for a given genotype, pods with more beans have slightly lighter beans. There would therefore probably be competition between beans within pods, i.e. beans may lack space for expansion when the pod contains many beans. However, this effect seems to be different depending on the genotypes; it is stronger for IMC 97 or GF 23 for instance (Fig. 2). These results lead us to propose protocols aimed at normalizing the phenotypic values of the genetic material. In order to obtain a reliable estimate of the bean weight, the following is proposed: i) either to use beans obtained from manual pollination to saturate the pods with beans, or ii) to systematically use the number beans in the pods as a covariable. By standardizing this measurement of bean weight, it will be easier to compare results from different laboratories.

The number of beans per pod depends on the number of ovules per ovary and the level of pollination of the flowers<sup>6</sup>. The number of ovules per ovary is a very heritable trait<sup>3</sup>; on the other hand, effective pollination depends on several factors: the self (in) compatibility of plant material<sup>17</sup>, the availability and intensity of pollinating insects traffic and therefore many environmental factors can have effects on the number of beans per pod<sup>9,18</sup>.

The average weight of the beans is not the same within a pod; the beans in the apical zone have significantly lighter beans than the beans in the other two zones (peduncular and median); it is the first time that this effect was studied and highlighted. As a result, resource partitioning is apparently not homogeneous within the pod, beans from areas further from the peduncle being smaller. The beans of clones with longer fruiting cycle durations are larger, i.e. the longer it takes for pods to mature, the larger their beans are. It was already observed for other fruits: the fruits with longer fruiting cycle durations are often larger<sup>19,20</sup>. In *T. cacao*, although fruit development is dependent on climatic factors such as temperature<sup>21</sup>, an important genetic variation on the fruiting cycle duration was already found<sup>22,23</sup>. It was shown that the longer the fruiting cycle duration, the heavier the beans are. However, with a longer fruiting cycle, the greater the risk exposure of the pods to diseases such as black pod disease<sup>24</sup> or frosty pod rot<sup>25</sup>. A trade-off has therefore to be found depending on the cultivation constraints of the areas.

The average weight of a bean is therefore a very heritable trait, but this trait is also dependent on several other characteristics: length of the fruiting cycle, number of beans per pod, location of the bean in the pods. The clones with Criollo ancestry had heavier beans (ICS 60, CCN 51). The length of the fruiting cycle of clones should be systematically characterized as it is a very heritable trait and explains several other traits of interest.

Received: 19 December 2019; Accepted: 29 May 2020;

Published online: 18 June 2020

## References

1. Motamayor, J. C. *et al.* Cacao domestication I: the origin of the cacao cultivated by the Mayas. *Heredity* **89**, 308–386 (2002).
2. Niemenak, N. *et al.* Phenological growth stages of cacao plants (*Theobroma* sp.): Codification and description according to the BBCH scale. *Annals of Applied Biology* **156**, 13–24 (2010).
3. Cilas, C., Machado, R. & Motamayor, J. C. Relations between several traits linked to sexual plant reproduction in *Theobroma cacao* L.: Number of ovules per ovary, number of seeds per pod, and seed weight. *Tree Genetics and Genomes* **6**, 219–226 (2010).



4. Ndoumbé, M., Bieysse, D. & Cilas, C. Multi-trait selection in a diallel crossing scheme of cocoa. *Plant Breeding* **120**, 365–367 (2001).
5. Tah, G. M. *et al.* Variation in yield over time in a cacao factorial mating design: changes in heritability and longitudinal data analyses over 13 consecutive years. *Euphytica* **215**, 106 (2019).
6. Falque, M., Vincent, A., Vaissière, B. E. & Eskes, A. B. Effect of pollination intensity on fruit and seed set in cacao (*Theobroma cacao* L.). *Sexual Plant Reproduction* **8**, 354–360 (1995).
7. Nyassé, S. *et al.* Integrated management of Phytophthora diseases on cocoa (*Theobroma cacao* L.): Impact of plant breeding on pod rot incidence. *Crop Protection* **26**, 40–45 (2007).
8. Sounigo, O. *et al.* Evaluation of resistance of *Theobroma cacao* L. to mirids in Côte d'Ivoire: results of comparative progeny trials. *Crop Protection* **22**, 615–621 (2003).
9. De Reffye, P. H., Parvais, J.-P., Mossu, G. & Lucas, P. Influence des aléas de la pollinisation sur les rendements du cacaoyer. *Café Cacao Thé* **22**, 251–274 (1978).
10. Cilas, C., Duchemin, C. & Lotodé, R. L'amélioration génétique de la qualité du cacao: étude de la granulométrie. *Café Cacao Thé* **33**, 3–8 (1989).
11. Engel, J. M. M. A systematic description of cacao clones. III. *Relationships between clones, between characteristics and some consequences for the cacao breeding*. *Euphytica* **32**, 719–733 (1983).
12. Bekele, F. L., Bekele, I., Butler, D. R. & Bidaise, G. G. Patterns of morphological variation in a sample of cacao (*Theobroma cacao* L.) germplasm from the International Cocoa Genebank, Trinidad. *Genetic Resources and Crop Evolution* **53**, 933–948 (2006).
13. Motamayor, J. C. & Lachenaud, P. e Mota, J. W. D. S., Loo, R., Kuhn, D. N., Brown, J. S. & Schnell, R. J. Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L.). *PloS one* **3**(10), e3311 (2008).
14. Agresti, A. & Coull, B. A. Approximate is better than "Exact" for interval estimation of binomial proportions. *The American Statistician* **52**, 119–126 (1998).
15. SAS Institute Inc. Base SAS® 9.3 Procedures Guide. Cary, NC: SAS Institute Inc (2011).
16. Enriquez, G. & Soria, J. V. The variability of certain bean characteristics of cacao (*Theobroma cacao* L.). *Euphytica* **17**, 114–20 (1968).
17. da Silva, M. R. *et al.* Genome-wide association mapping of sexual incompatibility genes in cacao (*Theobroma cacao* L.). *Tree Genetics and Genomes* **12**, 62 (2016).
18. Groeneveld, J. H., Tschardtke, T., Moser, G. & Clough, Y. Experimental evidence for stronger cacao yield limitation by pollination than by plant resources. *Perspectives in Plant Ecology, Evolution and Systematics* **12**, 183–191 (2010).
19. Jullien, A., Malézieux, E., Michaux-Ferrière, N., Chillet, M. & Ney, B. Within-bunch variability in banana fruit weight: importance of developmental lag between fruits. *Annals of Botany* **87**, 101–108 (2001).
20. Zhiguo, J., Yousheng, D. & Zhiqiang, J. Combinations of GA3 and AVG delay fruit maturation, increase fruit size and improve storage life of 'Feicheng' peaches. *J. Hort. Sci. and Biotech.* **74**, 579–583 (1999).
21. Daymond, A. J. & Hadley, P. Differential effects of temperature on fruit development and bean quality of contrasting genotypes of cacao (*Theobroma cacao*). *Annals of Applied Biology* **153**, 75–185 (2008).
22. Berry, D. & Cilas, C. Etude génétique de la réaction à la pourriture brune des cabosses chez des cacaoyers (*Theobroma cacao* L.) issus d'un plan de croisement diallele. *Agronomie* **14**, 599–609 (1994).
23. Ten Hoopen, G. M., Deberdt, P., Mbenoum, M. & Cilas, C. Modelling cacao pod growth: implications for disease control. *Annals of Applied Biology* **160**, 260–272 (2012).
24. Efombagn, M. I. B. *et al.* Effect of fruiting traits on the field resistance of cacao (*Theobroma cacao* L.) clones to Phytophthora megakarya. *Journal of Phytopathology* **152**, 557–562 (2004).
25. Jaimes, Y. *et al.* Factors affecting the dynamics of frosty pod rot in the main cocoa areas of Santander State, Colombia. *Plant Disease* **103**, 1665–1673 (2019).

## Acknowledgements

This study was part of the research program carried out in partnership with the research centre of CIRAD in Kourou (French Guiana) and the Research Unit « Bioagresseurs » of CIRAD. It was supported by CIRAD and particularly by the training service for the involvement of Fabien Doaré. We acknowledge the training service of CIRAD for the financial support of this work.

## Author contributions

F.D. and C.C. conceived the study. F.D. did the field observations. F.R. and C.C. carried out the data analyses. C.C. and F.D. wrote the paper. All authors reviewed and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to C.C.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020